

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2004, 11:12:16 ; Search time 1701.11 Seconds
(without alignments)
2142.114 Million cell updates/sec

Title: US-09-529-397C-25

Perfect score: 100

Sequence: 1 gggaguggaggaaucaucg.....uagacagcaagcuucg 100

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 1821965908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gss1:*

9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.6	34.6	642	BE147147	PM2-HT022
2	31.4	31.4	749	AG557770	MUS MUSCU
3	31.2	31.2	442	AZ955882	2M022110
4	30.8	30.8	696	B21536	F21F23-T7 I
5	30.6	30.6	461	CB736914	AMGNNUC:M
6	30.6	30.6	1074	BE371824	601217621
7	30.4	30.4	917	AG069018	Pan trogl
8	30.0	30.0	725	CK016754	AGENCOURT
9	29.8	29.8	528	AQ108017	CIT-HSP-2
10	29.6	29.6	578	CN553623	tae26g08
11	29.6	29.6	617	BG38990	Gm01_01e0
12	29.6	29.6	626	CN623811	tae59d12
13	29.6	29.6	717	CN553289	tae26g08
14	29.6	29.6	744	CN226782	RJB004H01
15	29.6	29.6	745	CO421969	GGEZHT100
16	29.6	29.6	852	B0209960	603949556
17	29.6	29.6	877	B0399017	603534483
18	29.2	29.2	254	BM401540	JH3C07R S
19	29.2	29.2	343	BG630802	cc-esf1cL
20	29.2	29.2	762	CF446768	EST683113
21	29.0	29.0	715	BZ944697	CH240_121
22	28.8	28.8	339	CN464219	7866_1 AF
23	28.8	28.8	532	CA249234	SCSBFL110
24	28.8	28.8	605	CE114143	tigr-gss-

25	28.8	28.8	739	8	CC115139	NDL.81L16
26	28.6	28.6	472	8	AQ955733	LERAF67TR
27	28.6	28.6	513	5	BP051024	BP051024
28	28.6	28.6	559	6	CB094897	h276e08.b
29	28.6	28.6	561	8	BZ891642	CH240_287
30	28.4	28.4	580	8	EZ201271	CH230-351
31	28.4	28.4	607	6	CA892404	B0172B04-
32	28.4	28.4	629	6	CA890800	B0161F04-
33	28.4	28.4	629	6	CA892977	B0175H04-
34	28.4	28.4	743	9	AG525698	MUS MUSCU
35	28.4	28.4	756	4	BG393101	602411369
36	28.4	28.4	818	5	B0708057	UI-M-PR0-
37	28.4	28.4	832	9	CC491386	CH240_325
38	28.4	28.4	835	6	CB686108	CB686108
39	28.2	28.2	496	5	BQ531573	APX2_5 F
40	28.2	28.2	521	5	BQ531501	APX3_35
41	28.2	28.2	521	7	CF791691	880031 MA
42	28.2	28.2	559	4	BG649176	EMI_77_E0
43	28.2	28.2	598	2	BE363373	WS1_62_B0
44	28.2	28.2	609	2	BE596636	P11_58_G0
45	28.2	28.2	625	6	CA267762	SCJLR207

ALIGNMENTS

RESULT 1
BE147147
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

PM2-HT0224-291099-002-c09 HT0224 Homo sapiens cdNA, mRNA sequence.
BE147147
BE147147.1 GI:8609871
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 642)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2002663
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=PM2-HT0224-291
099-002-c09&tl3=1999-10-29&tl4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 415.
Location/Qualifiers
1..642
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev stage="Adult"
/clone.lib="HT0224"
/note=Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 34.6%; Score 34.6; DB 2; Length 642;
Best Local Similarity 37.0%; Pred. No. 0.68; 29; Indels 0; Gaps 0;
Matches 30; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

QY 12 AAUUAUCGAGCAUUGUGCAGCUCGUCUUCUUAACCAACCAUUAUAAUUGGUUUAAG 71

DB 150 AATAAATTGAGGCTGTCTCAGATACCTTTTGGTTTGCAACAATAATAAATGGTTTAC 209

QY 72 CAUAUAGCCUAGGCAAGCA 92

DB 210 AATTTCGTTTACAACAACAA 230

RESULT 2

AG557770/c 749 bp DNA linear GSS 05-JUN-2004
LOCUS Mus musculus molossinus DNA, clone:MSMG01-476B14.T7, genomic survey
DEFINITION sequence.

ACCESSION AG557770

VERSION AG557770.1 GI:48318468

KEYWORDS GSS.

SOURCE Mus musculus molossinus

ORGANISM Mus musculus molossinus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

AUTHORS BAC end Sequences of Library MSMG01

TITLE Unpublished

JOURNAL (bases 1 to 749)

REFERENCE 2 Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.

AUTHORS Direct Submission

TITLE Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

JOURNAL and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/;

Tel:81-45-503-9111, Fax:81-45-503-9170)

Clones are derived from the mouse BAC library MSMG01. For BAC

library availability, please contact Kuniya Abe (abe@rtc.riken.jp).

Teukuba Institute, Bio Resource Center,

The Institute of Physical and Chemical Research (RIKEN) 3-1-1

Koyadai, Tsukuba, 305-0074 Japan

phone: 81-298-36-9189, fax: 81-298-36-9199

e-mail: abe@rtc.riken.jp

PRIMERS

Sequencing : T7

LIBRARY : pBAC3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1. .749

/organism="Mus musculus molossinus"

/mol_type="genomic DNA"

/sub_species="molossinus"

/db_xref="taxon:57486"

/clone="MSMG01-476B14.T7"

/sex="male"

/tissue="mixture of kidney and spleen"

/clone_lib="MSMG01 Mouse Male BAC Library"

ORIGIN

Query Match 31.4%; Score 31.4; DB 9; Length 749;
Best Local Similarity 51.0%; Pred. No. 8.6;
Matches 25; Conservative 13; Mismatches 11; Indels 0; Gaps 0;

QY 41 UUCUUCACCAACCAUUAUAAUUGGUUUUAGCAUAGCCUUAAGCAG 89
DB 539 TTCTTAAACCACTTATAAAATCTTCTAGCATATGCAATAGTGTCTG 491

RESULT 3

AG555882

LOCUS 442 bp DNA linear GSS 27-APR-2001

DEFINITION 2M0222I110F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

clone UUGC2M0222I10 F, Genomic survey sequence.

ACCESSION AG555882

VERSION AG555882.1 GI:13827109

KEYWORDS GSS.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 442)

AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,

Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,

Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von

Niederhausern,A. and Wright,D. Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

COMMENT Contact: Robert B. Weiss

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0222 row: I column: 10

Seq primer: CGTTGTAAACGACGCGCCACT

Class: plasmid ends

High quality sequence stop: 442.

Location/Qualifiers

1. 442

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clones="UUGC2M0222I10"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/note="Vector: FWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWP42 (gi|4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

ORIGIN

Query Match 31.2%; Score 31.2; DB 8; Length 442;
Best Local Similarity 32.6%; Pred. No. 9.4;
Matches 30; Conservative 24; Mismatches 38; Indels 0; Gaps 0;

```

QY 7 GGAGGAUUCAGGAGGAUUGCAGCUCUCCUUCUCAAACCGAUUAAUUGGU 66
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
272 GCAGGCTTTCATCCCGACATCTGATTCGATGCTTCATGATAAAATCTACAAATTGAG 331
QY 67 UUUAGCAUUGCCUAGCAGCAGCAGCUUCU 98
Db :::::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
332 TTTTGTGTTTGTGTTTTCGAGACCGAGTTTCT 363

RESULT 4
B21536
LOCUS
DEFINITION F21F23-T7 IGF Arabidopsis thaliana genomic clone F21F23, genomic
survey sequence.
ACCESSION B21536
VERSION B21536.1 GI:2396590
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 696)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F21F23-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 129
High quality sequence stop: 143.

FEATURES
source
1..696
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F21F23"
/sex="hermaphrodite"
/clone_lib="IGP"
/notes="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 30.8%; Score 30.8; DB 8; Length 696;
Best Local Similarity 43.9%; Pred. No. 14;
Matches 29; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 9 AGGAUUCAGGAGGAUUGCAGCUCUCCUUCUCAAACCGAUUAAUUGGUU 68
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
306 AGAAATATCATCGATCAAGTAGAAGGTGTGTCCTTGACACCATATAGATGTGTC 365
QY 69 UAGCAU 74
Db ::|||:
366 TACCAI 371

RESULT 5
CB736914/c
LOCUS
DEFINITION AMGNNUC-MRBE4-00015-C9-A rat brain E15 (10375) Rattus norvegicus
cDNA clone mrbe4-00015-c9 5', mRNA sequence.
ACCESSION CB736914
VERSION CB736914.1 GI:29804127

QY 7 GGAGGAUUCAGGAGGAUUGCAGCUCUCCUUCUCAAACCGAUUAAUUGGU 66
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
272 GCAGGCTTTCATCCCGACATCTGATTCGATGCTTCATGATAAAATCTACAAATTGAG 331
QY 67 UUUAGCAUUGCCUAGCAGCAGCAGCUUCU 98
Db :::::||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
332 TTTTGTGTTTGTGTTTTCGAGACCGAGTTTCT 363

RESULT 4
B21536
LOCUS
DEFINITION F21F23-T7 IGF Arabidopsis thaliana genomic clone F21F23, genomic
survey sequence.
ACCESSION B21536
VERSION B21536.1 GI:2396590
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 696)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: F21F23-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 129
High quality sequence stop: 143.

FEATURES
source
1..696
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="F21F23"
/sex="hermaphrodite"
/clone_lib="IGP"
/notes="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
Produced by Thomas Altmann"

ORIGIN
Query Match 30.8%; Score 30.8; DB 8; Length 696;
Best Local Similarity 43.9%; Pred. No. 14;
Matches 29; Conservative 15; Mismatches 22; Indels 0; Gaps 0;

QY 9 AGGAUUCAGGAGGAUUGCAGCUCUCCUUCUCAAACCGAUUAAUUGGUU 68
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
306 AGAAATATCATCGATCAAGTAGAAGGTGTGTCCTTGACACCATATAGATGTGTC 365
QY 69 UAGCAU 74
Db ::|||:
366 TACCAI 371

RESULT 5
CB736914/c
LOCUS
DEFINITION AMGNNUC-MRBE4-00015-C9-A rat brain E15 (10375) Rattus norvegicus
cDNA clone mrbe4-00015-c9 5', mRNA sequence.
ACCESSION CB736914
VERSION CB736914.1 GI:29804127

```

```

KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 461)
AUTHORS Angen EST Program.
TITLE Angen Rat EST Program
JOURNAL Unpublished (2003)
COMMENT Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00015 row: C column: 9.
FEATURES
source
1..461
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="mrbe4-00015-c9"
/tissue_type="brain E15"
/clone_lib="rat brain E15 (10375)"
/notes="Vector: pBCB; Site_1: BstXI; Site_2: NotI; rat
brain E15"

ORIGIN
Query Match 30.6%; Score 30.6; DB 6; Length 461;
Best Local Similarity 36.7%; Pred. No. 15;
Matches 29; Conservative 19; Mismatches 31; Indels 0; Gaps 0;

QY 8 GAGGAUUCAGGAGGAUUGCAGCUCUCCUUCUCAAACCGAUUAAUUGGU 67
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
180 GAGGACTTCATCAGTAGATGTCGCGATAAATTATAGTCTTGACAGATTATGGTGAGTT 121
QY 68 UUAGCAUUGCCUAGCGCA 86
Db ::|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
120 TGAGCAGTTGCATTAGCAA 102

RESULT 6
BE371824
LOCUS
DEFINITION S01217621F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3586661 5',
mRNA sequence.
ACCESSION BE371824
VERSION BE371824.1 GI:9317291
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1074)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM8748 row: j column: 06
High quality sequence stop: 380.
FEATURES
source
1..1074
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"

```

```

/clone="IMAGE:3586661"
/tissue_type="spontaneous tumor, metastatic to mammary.
Stem cell origin."
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Lu29"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match 30.6%; Score 30.6; DB 2; Length 1074;
Best Local Similarity 44.2%; Pred. No. 17;
Matches 34; Conservative 14; Mismatches 29; Indels 0; Gaps 0;

QY 19 CGAGCAUAGUGGACUCCUUCUCAAACCAAGUAAUUGUUUAGCAUAGC 78
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 CCAGGATTCAGACAGCCCTGCTCCCTGACCCCACTTATTAGATCTGTGAGCAATAC 339

QY 79 CUUAGCGACAGCAAGCU 95
Db |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
340 CATGGGACTTGATGCT 356

RESULT 7
AG069018/c
LOCUS AG069018 917 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-059H17.R, genomic survey sequence.
ACCESSION AG069018
VERSION AG069018.1 GI:16620820
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 917)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
E-mail:chimpses@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
Location/Qualifiers
1..917
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-059H17.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match 30.4%; Score 30.4; DB 9; Length 917;
Best Local Similarity 39.1%; Pred. No. 20;
Matches 25; Conservative 19; Mismatches 21; Indels 0; Gaps 0;

```

```

QY 22 GGCAUAGUGGACUCCUUCUCAAACCAAGUAAUUGUUUAGCAUAGCCUU 81
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 GGTATCTGTCGTCCTCCCTCTCTACGCTGGGTTATAATGATTTAAATTACCTT 307

QY 82 AGCG 85
Db |||
306 CCCG 303

RESULT 8
CK016754/c
LOCUS CK016754 725 bp mRNA linear EST 26-NOV-2003
DEFINITION AGENCOURT 16544280 NIH_ZGC_10 Danio rerio cDNA clone IMAGE:7043617
5', mRNA sequence.
ACCESSION CK016754
VERSION CK016754.1 GI:38541867
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
REFERENCE 1
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Len Zon, Harvard
cDNA Library Preparation: Open Biosystems
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14803 Row: i column: 23
High quality sequence stop: 594.
FEATURES
source
Location/Qualifiers
1..725
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:7043617"
/tissue_type="whole body"
/lab_host="DH10B"
/clone_lib="NIH_ZGC_10"
/notes="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;
Bulk tissue was collected from a whole adult individual
from the Tuebingen strain. 1st strand cDNA was primed with
a Not I - oligo(dT) primer, double-stranded cDNA was
cloned into the Not I and EcoRV sites of pExpress-1.
Library was size-selected for >1 kb fragments. A
normalized version of this library is also available
(NIH_ZGC 7). Library was constructed by Open Biosystems
(Huntsville, AL)."
```

ORIGIN

```

Query Match 30.0%; Score 30; DB 7; Length 725;
Best Local Similarity 44.3%; Pred. No. 26;
Matches 31; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

QY 2 GGAGUGAGGAUUCUCCGAGCAUUGUCCGUCUCCUCAAACCAAGUAAUUG 61
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 GGAGTGAAGATTTCAGACAGACATTGTGGGCATCGTAGTCTTTGAAGCATATATA 604

QY 62 UUGUUUUUAG 71
Db |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
603 TAAGTAGTAG 594

```

```

RESULT 9
AQ108017/c
LOCUS
DEFINITION
  CIT-HSP-2374K6.TF CIT-HSP Homo sapiens genomic clone 2374K6,
  genomic survey sequence.
ACCESSION
  AQ108017
KEYWORDS
  GSS.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 528)
  Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
  Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
  Venter,J.C.
  Use of a random human BAC End Sequence Database for Sequence-Ready
  Map Building
  Unpublished (1998)
JOURNAL
  Other GSSs: CIT-HSP-2374K6.TR
COMMENT
  Contact: Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: mdadams@tigr.org
  Clones are available from Research Genetics (info@resgen.com). BAC
  end search page:
  http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
  Seq primer: M13-21
  Class: BAC ends.
FEATURES
  source
  1..528
  Location/Qualifiers
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /clone="2374K6"
    /sex="Male"
    /cell_type="Sperm"
    /clone_lib="CIT-HSP"
    /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
    HindIII"
ORIGIN
  Query Match 29.8%; Score 29.8; DB 8; Length 528;
  Best Local Similarity 36.9%; Pred. No. 29;
  Matches 24; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
  QY 13 AUUUAUGGAGGCAUAGUCGACUCCUCCUUAACCAACCAUUAUUAUUGGUUUUAGC 72
  DB 411 ATTATACAGGTATATTGACTGCGTTTCTTGGACCAAAATACAAATGAATTTAT 352
  QY 73 AUAUG 77
  DB 351 ACATG 347
  CN553623
  tae26g08.v1 Hydra EST Darmstadt I Hydra magnipapillata cDNA 5'
  similar to TR.Q9VUZ2 Q9VUZ2 CG5656 PROTEIN. ;, mRNA sequence.
ACCESSION
  CN553623
VERSION
  CN553623.1 GI:46962927
SOURCE
  EST.
  Hydra magnipapillata
  Hydra magnipapillata
  Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
  Hydridae; Hydra.
REFERENCE
  1 (bases 1 to 578)
  Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
  Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,

```

```

  Hillier L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
  Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarisvili,R.,
  Maguire,D., Kennedy,S., Waterston,R. and Wilson,R.
  WashU Hydra EST Project
  Unpublished (2002)
  Other ESTs: tae26g08.x1
  Contact: Hans Bode
  WashU Hydra EST Project
  Washington University School of Medicine
  444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@wustl.edu
  Library was constructed by Corina Guder / GATC Konstanz, Germany
  Library materials provided by Thomas Holstein / Molecular Cell
  Biology, TUD, Darmstadt DNA sequencing by: Washington University
  Genome Sequencing Center For information on obtaining a clone
  please contact: Hans Bode (hnbode@uci.edu)
  Seq primer: -40UP
  High quality sequence stop: 565.
FEATURES
  source
  1..578
  Location/Qualifiers
    /organism="Hydra magnipapillata"
    /mol_type="mRNA"
    /strain="sf-1 mutant of strain 105"
    /db_xref="taxon:6085"
    /lab_host="TransforMax EC100 (Epicentre), T1 Phage
    resistant electrocompetent cells"
    /clone_lib="Hydra EST Darmstadt I"
    /note="Vector: pBluescript II SK (+); Site_1: NotI;
    Site_2: EcoRI"
ORIGIN
  Query Match 29.6%; Score 29.6; DB 7; Length 578;
  Best Local Similarity 40.8%; Pred. No. 34;
  Matches 31; Conservative 16; Mismatches 29; Indels 0; Gaps 0;
  QY 21 AGCAUAUGGACUCCUCCUCCUUAACCAACCAUUAUUAUUGGUUUUAGCAUAGCCU 80
  DB 453 AAGGATACGGGCACCTTGTGACCCCAATGAACAAAGTTGTGAATTCATCATTTGAGT 394
  QY 81 UAGCGACAGCAAGCUU 96
  DB 393 CAACGAGATCAAGTTT 378
  B838990/c
  LOCUS
  DEFINITION
    B838990 617 bp mRNA linear EST 25-MAY-2001
    Glycine max cDNA Clone Gm01_01e04, mRNA sequence.
  ACCESSION
    B838990
  VERSION
    B838990.1 GI:14205312
  KEYWORDS
    EST.
  SOURCE
    Glycine max (soybean)
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
    Glycine.
  REFERENCE
    1 (bases 1 to 617)
  AUTHORS
    Singh,J.A., Robert,L.S., Lu,B., Zhu,L., De Moors,A., Couroux,P.,
    Harris,L.J., Hattori,J.I., Quillet,T., Spratt,D. and Tinker,N.A.
    Expressed Sequence Tags from Cold-Stressed Glycine max (soybean)
    Leaves
    Unpublished (2001)
  JOURNAL
    Contact: Singh,J.A.
  COMMENT
    Eastern Cereal and Oilseed Research Centre
    Agriculture and Agri-food Canada
    KW Neatby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
    0C6, Canada
    Tel: (613) 759-1662
    Fax: (613) 759-1701

```



```

QY      81  UAGCGACACGAAGCUU 96
      |||||
Db      526  CAACGAGATCAAGTTT 541

RESULT 14
CN226782      744 bp  mRNA  linear  EST 09-APR-2004
LOCUS      RJB004H01.ab1 RJ-testis Gallus gallus cDNA 5', mRNA sequence.
DEFINITION  CN226782
ACCESSION  CN226782
VERSION    CN226782.1 GI:46330526
KEYWORDS  EST
SOURCE    Gallus gallus (chicken)
ORGANISM  Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 744)
AUTHORS   Savolainen, P., Fitzsimmons, C.J., Arvestad, L., Andersson, L. and
           Lundberg, J.
TITLE     EST analysis of brain and testis cDNA libraries from White leghorn
           and Red Jungle Fowl
JOURNAL   Unpublished (2004)
COMMENT   Department of Biotechnology
           Royal Institute of Technology, KTH
           SE-106 91 Stockholm, SWEDEN
           Tel: +46 (0)8 5537 8481
           Fax: +46 (0)8 5537 8335
           Email: Peter.Savolainen@biotech.kth.se
           Seq primer: M13 reverse primer.

FEATURES             Location/Qualifiers
     source            1..744
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="Red junglefowl"
                     /db_xref="taxon:9031"
                     /sex="male"
                     /lab_host="ElectroMAX DH10B (Invitrogen)"
                     /clone_lib="RJTestis"
                     /note="Organ: testis; Vector: pSPORT-1; Site 1: Hind III;
                           Site 2: EcoRI; The cDNA libraries were created with the
                           Superscript Plasmid System (Invitrogen)."
```

Query Match 29.6%; Score 29.6; DB 7; Length 744;
 Best Local Similarity 55.8%; Pred. No. 36;
 Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

```

QY      1  GGGAGUGGAGGAUAUCAGGACGAUAGUGCGACUCGUCUCCUCAAACC 52
      |||||
Db      491  GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGGTCTCCCATCGCACC 542

RESULT 15
CO421969      745 bp  mRNA  linear  EST 02-JUL-2004
LOCUS      GGEZHT1005A10.g HT1 Gallus gallus cDNA clone GGEZHT1005A10, mRNA
DEFINITION  sequence.
ACCESSION  CO421969
VERSION    CO421969.1 GI:49638217
KEYWORDS  EST
SOURCE    Gallus gallus (chicken)
ORGANISM  Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 745)
AUTHORS   Silva, C.S., Jorge, E.C., Patricio, M., Ledur, M.C. and Coutinho, L.L.
TITLE     Discovery of new genes expressed in the chicken pituitary and
           hypothalamus
JOURNAL   Unpublished (2004)
COMMENT   Contact: Clárisa S. Silva
```

Query Match 29.6%; Score 29.6; DB 7; Length 745;
 Best Local Similarity 55.8%; Pred. No. 36;
 Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

```

QY      1  GGGAGUGGAGGAUAUCAGGACGAUAGUGCGACUCGUCUCCUCAAACC 52
      |||||
Db      254  GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGGTCTCCCATCGCACC 305

Search completed: November 30, 2004, 12:52:13
Job time : 1709.11 secs
```

Query Match 29.6%; Score 29.6; DB 7; Length 745;
 Best Local Similarity 55.8%; Pred. No. 36;
 Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

```

QY      1  GGGAGUGGAGGAUAUCAGGACGAUAGUGCGACUCGUCUCCUCAAACC 52
      |||||
Db      254  GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGGTCTCCCATCGCACC 305
```

Search completed: November 30, 2004, 12:52:13
 Job time : 1709.11 secs

```

Laboratory of Animal Biotechnology, Dep. of Animal Production
ESALQ - University of Sao Paulo
Av. Pádua Dias, 11, Piracicaba, SP, 13418-900, Brazil
Tel: 55 19 3429 4434
Fax: 55 19 3429 4285
Email: cssilva@esalq.usp.br and llcoutine@esalq.usp.br
PCR Primers
BACKWARD: T7.

FEATURES             Location/Qualifiers
     source            1..745
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9031"
                     /clone="GGEZHT1005A10"
                     /tissue_type="pituitary and hypothalamus"
                     /dev_stage="21 days old"
                     /lab_host="DH10B"
                     /clone_lib="HT1"
                     /notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
                           cDNA library was constructed with the SuperScript Plasmid
                           System with Gateway Technology kit (Invitrogen) following a
                           manufacturer's protocols. Plasmid DNA was purified using a
                           modified alkaline lysis method. Sequencing reactions were
                           conducted using the kit Big Dye Terminator Cycle
                           Sequencing Ready Reaction (Applied Biosystems) according
                           to the manufacturer's recommendations. Clones were
                           sequenced by the 5' end with T7 primer. Sequencing
                           reactions were analyzed on ABI Prism 3100 Genetic Analyzer
                           (Applied Biosystems). The quality and clustering of the
                           ESTs were analyzed using the softwares Phred/Cap3. Only
                           EST sequences with Phred quality greater than 20 and at
                           least 150 bp were considered for clustering."
```

Query Match 29.6%; Score 29.6; DB 7; Length 745;
 Best Local Similarity 55.8%; Pred. No. 36;
 Matches 29; Conservative 9; Mismatches 14; Indels 0; Gaps 0;

```

QY      1  GGGAGUGGAGGAUAUCAGGACGAUAGUGCGACUCGUCUCCUCAAACC 52
      |||||
Db      254  GGGAGTTTAAGGCTTCATCGAGGCAAAATGGGAGCTCGGTCTCCCATCGCACC 305
```

Search completed: November 30, 2004, 12:52:13
 Job time : 1709.11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2004, 11:12:15 ; Search time 1093.89 Seconds
(without alignments)
4323.081 Million cell updates/sec

Title: US-09-529-397C-25
Perfect score: 100
Sequence: 1 gggaguggaggaaucaucg.....uagcagacagcaagcuucgc 100

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_uni.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32.2	32.2	199415	10 AC084053	AC084053 Mus Muscu
2	31.2	31.2	201197	10 AC125279	AC125279 Mus muscu
3	30.8	30.8	187606	2 AC110847	AC110847 Rattus no
4	30.8	30.8	221981	2 AC103165	AC103165 Rattus no
5	30.8	30.8	231814	2 AC114165	AC114165 Rattus no
6	30.8	30.8	233823	2 AC110697	AC110697 Rattus no
7	30.8	30.8	238172	2 AC099144	AC099144 Rattus no
8	30.8	30.8	250740	2 AC129051	AC129051 Rattus no
9	30.6	30.6	748	6 CQ451733	CQ451733 Sequence
10	30.6	30.6	15776	6 CQ363795	CQ363795 Sequence
11	30.6	30.6	110000	1 AE017283_15	Continuation (16 o
12	30.4	30.4	217	4 AF522912	AF522912 Tachyglos
13	30.4	30.4	217	4 AF522914	AF522914 Tachyglos
14	30.4	30.4	217	4 AF522916	AF522916 Tachyglos
15	30.4	30.4	366	6 AR318058	AR318058 Sequence
16	30.4	30.4	3185	4 AY194920	AY194920 Tachyglos
17	30.4	30.4	46077	2 AC137289	AC137289 Rattus no
18	30.4	30.4	58143	2 AC137177	AC137177 Rattus no
19	30.4	30.4	207260	2 AC134220	AC134220 Rattus no

20	30.4	213033	2	AC131482	AC131482 Rattus no
21	30.4	218157	2	AC139588	AC139588 Rattus no
22	30.4	272400	2	AC110146	AC110146 Rattus no
23	30.4	283299	2	AC112576	AC112576 Rattus no
24	30.2	191004	2	CR536601	CR536601 Ornithorh
25	30.2	229522	2	AC109911	AC109911 Rattus no
26	30.2	235937	2	AC131223	AC131223 Rattus no
27	30.2	240755	2	AC095368	AC095368 Rattus no
28	30.0	213257	2	AC133102	AC133102 Mus muscu
29	29.8	4285	6	AR372457	AR372457 Sequence
30	29.8	4285	8	AF057708	AF057708 Populus b
31	29.8	184147	9	AC103805	AC103805 Homo sapi
32	29.8	191844	2	AC148836	AC148836 Pan trogl
33	29.8	202287	2	AC145887	AC145887 Pan trogl
34	29.8	215994	9	AC067941	AC067941 Homo sapi
35	29.8	223143	2	CR450730	CR450730 Danio rer
36	29.6	23393	2	AC018216	AC018216 Drosophil
37	29.6	50481	2	AC117541	AC117541 Mus muscu
38	29.6	79826	3	AC002443	AC002443 Drosophil
39	29.6	110000	2	AC110832_3	Continuation (4 of
40	29.6	110000	2	AC112373_04	Continuation (5 of
41	29.6	156805	3	AC093047	AC093047 Drosophil
42	29.6	182816	5	AC145913	AC145913 Gallus ga
43	29.6	192132	3	AC093440	AC093440 Drosophil
44	29.6	234627	2	AC106118	AC106118 Rattus no
45	29.6	237005	2	AC137343	AC137343 Rattus no

ALIGNMENTS

RESULT 1	AC084053	Mus Musculus Strain C57BL6/J chromosome 3 BAC, RP23-158F1, complete sequence.	199415 bp	DNA	linear	ROD 05-SEP-2002
AC084053	AC084053	AC084053.6	GI:22725947	HTG.		
LOCUS	AC084053					
DEFINITION	AC084053					
ACCESSION	AC084053					
VERSION	AC084053.6					
KEYWORDS	HTG.					
SOURCE	Mus musculus (house mouse)					
ORGANISM	Mus musculus					
REFERENCE	1	(Bases 1 to 199415)				
AUTHORS	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusiina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.					
TITLE	High Throughput Mouse Sequencing					
JOURNAL	Unpublished					
REFERENCE	2	(Bases 1 to 199415)				
AUTHORS	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusiina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.					
TITLE	Direct Submission					
JOURNAL	Submitted (12-OCT-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA					
REFERENCE	3	(Bases 1 to 199415)				
AUTHORS	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusiina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.					
TITLE	Direct Submission					
JOURNAL	Submitted (29-JUL-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA					
REFERENCE	4	(Bases 1 to 199415)				
AUTHORS	Han, J., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusiina, M., Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J., Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.					
TITLE	Direct Submission					
JOURNAL	Submitted (05-SEP-2002) Harvard Partners Center for Genetics and Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA 02139, USA					


```

repeat_region 23882..24118
/rpt_family="Lx4"
repeat_region complement(24228..24641)
/rpt_family="Lx6"
repeat_region complement(24821..25092)
/rpt_family="Lx3"
repeat_region 25183..25208
/rpt_family="TAAA)n"
repeat_region 25210..25277
/rpt_family="GAAA)n"
repeat_region complement(25288..25661)
/rpt_family="MTC"
repeat_region complement(26114..32348)
/rpt_family="LIF"

Query Match 32.2%; Score 32.2; DB 10; Length 199415;
Best Local Similarity 49.1%; Pred No. 8.5;
Matches 26; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY 37 CGUCUCCUCAAACAGUUAUAUUGGUUUUAGCAUUGCUUAGCGACAG 89
Db 63221 CTTATTCCTTAAACACAGTATATAAATCTTGTCATATGCAATAGTGCTG 63273

RESULT 2
AC125279/c
LOCUS Mus musculus BAC clone RP23-433F5 from 8, complete sequence.
DEFINITION AC125279
ACCESSION AC125279
VERSION AC125279.1 GI:21536173
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201197)
Trani, L. and Cotton, M.
The sequence of Mus musculus BAC clone RP23-433F5
Unpublished (2001)
Wilson, R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 201197)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 201197)
Wilson, R.
Direct Submission
Submitted (05-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@genome.wustl.edu
----- Summary Statistics
----- Center project name: M_BA04333P05
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by

restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tatenio in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES	Location/Qualifiers
source	1..201197 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="8" /map="8"
repeat_region	/clone="RP23-433F5" /clone_lib="RPCI-23" 171..557 /rpt_family="MaLR"
repeat_region	558..1352 /rpt_family="MaLR"
repeat_region	1353..1499 /rpt_family="Alu"
repeat_region	1500..1886 /rpt_family="MaLR"
repeat_region	1887..2101 /rpt_family="MaLR"
repeat_region	2103..2485 /rpt_family="L1"
repeat_region	2732..2918 /rpt_family="Alu"
repeat_region	2916..3337 /rpt_family="RMR6B"
repeat_region	5516..5712 /rpt_family="MaLR"
repeat_region	6071..6162 /rpt_family="ERVK"
repeat_region	6163..6556 /rpt_family="MaLR"
repeat_region	6557..6980 /rpt_family="ERVK"
repeat_region	7137..7220 /rpt_family="MIR"
repeat_region	7970..8262 /rpt_family="L1"
repeat_region	8734..9125 /rpt_family="L1"
repeat_region	9194..10473 /rpt_family="L1"
repeat_region	10485..10619 /rpt_family="MER1_type"
repeat_region	10573..10940 /rpt_family="ERVK"
repeat_region	10673..11051 /rpt_family="RMR17C"
repeat_region	11200..12140 /rpt_family="L1"
repeat_region	12674..12714 /rpt_family="L1"
repeat_region	12715..12907 /rpt_family="B2"
repeat_region	12908..13742 /rpt_family="L1"

14760. .15436
/rpt_family="L1"
repeat_region
16028. .16131
/rpt_family="MaLR"
repeat_region
17836. .18263
/rpt_family="L1"
repeat_region
18352. .18653
/rpt_family="ERVK"
repeat_region
18655. .19078
/rpt_family="ERVK"
repeat_region
19375. .19942
/rpt_family="L1"
repeat_region
21030. .21117
/rpt_family="MaLR"
repeat_region
22308. .22444
/rpt_family="MIR"
repeat_region
22775. .22988
/rpt_family="MER1_type"
repeat_region
22991. .23088
/rpt_family="L1"
repeat_region
23274. .23681
/rpt_family="MaLR"
repeat_region
24784. .24971
/rpt_family="B2"
repeat_region
25314. .25486
/rpt_family="MaLR"
repeat_region
26145. .26223
/rpt_family="RAL_RN"
repeat_region
26473. .26619
/rpt_family="Alu"
repeat_region
26817. .27001
/rpt_family="B2"
tRNA
complement(26924. .26996)
/product="tRNA-Ser"
/note="likely pseudogene (HMM Sc=39.59 / Sec struct
Sc=-11.21)"
repeat_region
27016. .27079
/rpt_family="Alu"
repeat_region
28465. .28604
/rpt_family="Alu"
repeat_region
28815. .29133
/rpt_family="MaLR"
repeat_region
29251. .29389
/rpt_family="MIR"
repeat_region
29546. .29687
/rpt_family="MaLR"
repeat_region
30593. .30737
/rpt_family="L2"
repeat_region
32073. .32437
/rpt_family="MaLR"
repeat_region
33137. .33276
/rpt_family="Alu"
repeat_region
33402. .33547
/rpt_family="Alu"
repeat_region
35207. .35290
/rpt_family="MIR"
repeat_region
35653. .35830
/rpt_family="MaLR"
repeat_region
35831. .36039
/rpt_family="B2"
repeat_region
36040. .36232
/rpt_family="MaLR"
repeat_region
36324. .36549
/rpt_family="B4"
repeat_region
36709. .36788
/rpt_family="B4"
repeat_region
36795. .36891
/rpt_family="ERV1"
repeat_region
36882. .37099
/rpt_family="MaLR"
repeat_region
39693. .39806
/rpt_family="MER1_type"
repeat_region
40903. .41102

/rpt_family="MER2_type"
41131. .41296
/rpt_family="MER2_type"
41519. .41626
/rpt_family="L2"
42585. .42730
/rpt_family="Alu"

Query Match 31.2% Score 31.2; DB 10; Length 201137;
Best local similarity 32.6%; Pred. No. 19;
Matches 30; Conservative 24; Mismatches 38; Indels 0; Gaps 0;

QY 7 GGAGAAUUCAGGAGCAUAGUGGACUCCGUCUCCUACACCAAGUUAUUAUUGU 66
Db 2584 GCAGCTTTTCATCCAGCATCTGATTCCTGAGTCTTCAATGATAAATCTAACATTTAG 2925
QY 67 UUVAGCAUUGCCUUAAGCAGCAGCAAGCUUCU 98
Db 2924 TTTTGTGTTTGTGTTTCGAGACAGGTTTCT 2893

RESULT 3
AC110847
LOCUS
DEFINITION Rattus norvegicus clone CH230-42M10, *** SEQUENCING IN PROGRESS
*** 13 unordered pieces.
AC110847
VERSION AC110847.6 GI:25006792
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 187606)
AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., C. Dederich, D., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, S., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Flopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Fu, L., L., Fuazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:238172.1

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: CH230-68F7
Center clone name: CH230-68F7
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 220106 bases at least Q40
Consensus quality: 222963 bases at least Q30
Consensus quality: 224727 bases at least Q20
Estimated insert size: 230633; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 238172: contig of 238172 bp in length.
Location/Qualifiers
1..238172
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-68F7"
1..1721
/note="wgs end extension
clone_end:Sp6"
8345_-8845
/note="clone boundary
clone_end:Sp6
site:EcoRI
end_sequence:BH351992"
234541..234906
/note="clone boundary
clone_end:T7"
site:EcoRI
end_sequence:BH351956"
236172..238172
/note="wgs end extension
clone_end:T7"

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature

ORIGIN
Query Match 30.8%; Score 30.8; DB 2; Length 238172;
Best Local Similarity 43.1%; Pred. No. 26;
Matches 25; Conservative 16; Mismatches 17; Indels 0; Gaps 0;
23 GCAUAGUCGACGCGGCUUCCUACCAACCAAGUUAUUAUUGGCUUAGCAUAGCCU 80


```

misc_feature      179234..181122
                  /note="wgs_contig"

ORIGIN
Query Match      30.8%; Score 30.8; DB 2; Length 250740;
Best Local Similarity 43.1%; Pred. No. 26;
Matches 25; Conservative 16; Mismatches 17; Indels 0; Gaps 0;

Qy      23   GCAUAGUGGACACUGGUCUUCUUCACCAACCAAGUUAUAAUUGUUUUAUGCAUAUGCCU 80
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
4082   GCATATTCGATTCCTCTTCCTCTTAACCAATTTGAACGCGTGTCTTCAAAAATCTCT 4025
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

```

[illegible]

```

ORIGIN
Query Match          30.6%; Score 30.6; DB 6; Length 748;
Best Local Similarity 46.8%; Pred. No. 20;
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;

Qy 18 UCGAGGCAUAUGUCGACUUCGUCUUCUUCUCAAACACAGUUAUAUAUUGUUUUUAGCAUAUG 77
Db 317 TTGGGGTCGACCTCGACGCCAGCGTCTCTCAGACCGTTGATGAACCGGTTGTAGGTCATG 258

Qy 78 CCTUAGCGACAGCAAGC 94
Db 257 CCCTGGGCACGGGAAGC 241

```

RESULT 10					
CQ363795	CQ363795	15776 bp	DNA	linear	PAT 23-JUN-2004
LOCUS	Sequence 78 from Patent WO0181581.				
DEFINITION	CQ363795				
ACCESSION	CQ363795				
VERSION	CQ363795.1	GI:41300489			
KEYWORDS	.				
SOURCE	Propionibacterium acnes				
ORGANISM	Propionibacterium acnes				
REFERENCE	Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Propionibacterineae; Propionibacteriaceae; Propionibacterium.				
AUTHORS	1 Skeiky,Y.A., Persing,D.H., Mitcham,J.L., Wang,S.S., Bhatia,A., L'Maisonneuve,J.F., Zhang,Y., Jen,S. and Carter,D.				
TITLE	Compositions and methods for the therapy and diagnosis of acne vulgaris				
JOURNAL	Patent: WO 0181581-A 78 01-NOV-2001;				
FEATURES	CORIXA CORPORATION (US)				
source	Location/Qualifiers				
	1..15776				
	/organism="Propionibacterium acnes"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:1747"				

ORIGIN

AE017283-23 2300001 2300283
Continuation (16 of 26) of AE017283 from base 1500001 (AE017283 Propionibacterium acnes)

	Query Match	30.5%	Score 30.6;	DB 1;	Length 110000;
	Best Local Similarity	46.8%;	Fred. No. 29;		
	Matches 36;	Conservative	12;	Mismatches 29;	Indels 0; Gaps 0;
Qy	18	UCGAGGCAUAGUCGACUCCGUCUCCUACAAACGAGUUAUAAUUGUUUUAAGCAUAG	77		
Db	34666	TTGCGGTCGACCTCGAGCGCAGCGTTCTTCAGACCGTGTGAGAAACGGITGTAGGTCATG	34722		
Qy	78	CCUAGCGCAGCAAGC	94		
Db	34726	CCCTGGGCACGGGAAGC	34742		

RESULT 12	
AF522912/c	
LOCUS	AF522912
DEFINITION	Tachyglossus aculeatus clone Tag7 type I interferon gene, partial cds.
ACCESSION	AF522912
VERSION	AF522912.1 GI:27451581
KEYWORDS	.
SOURCE	Tachyglossus aculeatus (Australian echidna)
ORGANISM	Tachyglossus aculeatus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Monotremata; Tachyglorissidae; Tachyglossus.
REFERENCE	1 (bases 1 to 217)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2004, 11:12:11 ; Search time 232.222 Seconds
(without alignments)
2260.516 Million cell updates/sec

Title: US-09-529-397C-25

Perfect score: 100

Sequence: 1 gggaguggggaaucaucg.....uagcagcagcaagcuucg 100

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseqn_23Sep04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	100.0	100	3	Aaz99048 RNA aptam
2	100	100.0	100	12	ADJ10063 Synthetic
3	100	100.0	100	12	ADJ10064 Synthetic
4	100	100.0	200	12	ADJ10065 Synthetic
5	90	90.0	90	3	Aaz99050 RNA aptam
6	84.4	84.4	98	12	ADJ10053 RNA aptam
7	80	80.0	80	3	Aaz99051 RNA aptam
8	50.6	50.6	147	12	ADJ10066 Synthetic
9	45	45.0	45	3	Aaz99076 Ras target
10	30.6	30.6	748	3	Aac75113 Human ORF
11	30.6	30.6	748	6	ABN24508 Human ORF
12	30.6	30.6	15776	8	Aaz99583 Propionib
13	30.6	30.6	15776	8	ACF64512 Propionib
14	30.4	30.4	366	9	ADA29321 DNA encod
15	30	30.0	98	3	Aaz99049 RNA aptam
16	29.8	29.8	4192	3	Aaz75942 Poplar fl
17	29.8	29.8	4285	4	Aaf85391 Nucleotid
18	29.8	29.8	4285	6	ABK88484 Poplar pr
19	29.8	29.8	4285	9	ACA62517 Poplar ho
20	29	29.0	29	3	Aaz99084 Ras target
21	29	29.0	29	12	ADJ10055 Synthetic

C 22	28.8	28.8	675	8	ACA52761
C 23	28.6	28.6	2000	6	ABZ15103 Arabidops
C 24	28.2	28.2	1161	5	ABV24321 Human pro
C 25	28.2	28.2	1663	2	AAX58401 Vicia sat
C 26	28	28.0	29	12	ADJ10058
C 27	28	28.0	29	12	ADJ10057
C 28	28	28.0	29	12	ADJ10056
C 29	28	28.0	110000	6	ABX08336 01
C 30	28	28.0	110000	12	ADJ25985 01
C 31	28	28.0	110000	12	ADN97989 01
C 32	28	28.0	110000	12	ADJ05281 01
C 33	27.6	27.6	732	4	AAI94947
C 34	27.6	27.6	732	8	ABT42823
C 35	27.4	27.4	3791	2	AAI63575
C 36	27	27.0	27	3	Aaz99087
C 37	27	27.0	29	12	ADJ10059
C 38	27	27.0	208	3	AAC17425
C 39	27	27.0	522	3	AAA69206
C 40	27	27.0	56506	3	AAA69168
C 41	26.8	26.8	415	3	AAC21986
C 42	26.8	26.8	1116	5	AAH51835
C 43	26.8	26.8	1116	12	ADH44700
C 44	26.6	26.6	48995	9	ADA02579
C 45	26.6	26.6	48995	10	ADB72317

ALIGNMENTS

RESULT 1

Aaz99048
ID Aaz99048 standard; RNA; 100 BP.

AC Aaz99048;

DT 21-JUN-2000 (first entry)

DE RNA aptamer #25 for binding Ras target protein.

XX Ras target protein; malignant tumour; signal transduction regulation;
KW cell proliferation; cell differentiation; aptamer; inflammation; ss.
XX Homo sapiens.

OS

PN WO200009684-A1.

PD 24-FEB-2000.

PF 13-AUG-1999; 99WO-JP004399.

PR 14-AUG-1998; 98JP-00242596.

PR 24-NOV-1998; 98JP-00333284.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Yokoyama S, Hirao I, Sakamoto K;

XX WPI; 2000-224330/19.

XX Nucleic acid e.g. RNA aptamer capable of binding specifically to Ras
PT target protein like Raf-1, useful in drug compositions to treat and
PT diagnose malignant tumors and inflammation.

XX Claim 6; Page 41; 59pp; Japanese.

XX The invention relates to novel nucleic acids which bind specifically to
CC the target protein of Ras, e.g. Raf-1. RNA aptamer (AAZ99048-239051)
CC based on these sequences are useful in the treatment and diagnosis of
CC malignant tumours and inflammation. The nucleic acids can be used to
CC formulate medicinal compositions that are useful in the treatment of
CC malignant tumours and inflammation and for disease diagnosis by binding
CC specifically to Ras target protein and regulating transmission of signal
CC causing proliferation or differentiation of cells

XX SQ Sequence 100 BP; 25 A; 22 C; 24 G; 0 T; 29 U; 0 Other;
 Query Match 100.0%; Score 100; DB 3; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.4e-25;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGUGAGGAAUUCAGCAGGCAUAGUGCGACUCCGUCUCCUCAAACCGAUUAA 60
 DB 1 GGGAGUGAGGAAUUCAGCAGGCAUAGUGCGACUCCGUCUCCUCAAACCGAUUAA 60

QY 61 AUUGUUUAGCAUAGCCUAGCGACAGCAAGCUCUCCG 100
 DB 61 AUUGUUUAGCAUAGCCUAGCGACAGCAAGCUCUCCG 100

RESULT 2
 ADJ10063
 ID ADJ10063 standard; DNA; 100 BP.
 XX AC ADJ10063;
 XX 17-JUN-2004 (first entry)
 XX Synthetic RNA 9A (100-mer).
 XX ss; unnatural base; 5'-substituted-2-oxo(1H) - pyridin-3-yl;
 XX photoreactive group; biotin; fluorescent molecule;
 XX 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.
 XX Synthetic.
 XX OS
 XX WO2004007713-A1.
 XX 22-JAN-2004.
 XX 28-FEB-2003; 2003WO-JP002342.
 XX 17-JUL-2002; 2002JP-00208568.
 XX (RIKE) RIKEN KK.
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX Hirao I, Yokoyama S, Hirao M, Mitsui T;
 XX WPI; 2004-122944/12.
 XX Nucleosides or nucleotides comprising novel base particularly 5-
 PT substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing
 PT functional nucleic acids as reagents or therapeutics in treating
 PT diseases.
 XX Disclosure; Fig 10; 78pp; Japanese.

XX This invention relates to nucleosides or nucleotides that a novel
 CC unnatural base, namely the 5'-substituted-2-oxo(1H) - pyridin-3-yl group
 CC as a base. Specifically, it refers to nucleic acid molecules integrated
 CC with a base substituted at the 5-position with either a photoreactive
 CC group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group,
 CC biotin or a derivative thereof or a fluorescent molecule selected from
 CC fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrhodamine or
 CC derivatives thereof. The present invention describes preparing nucleic
 CC acids by carrying out transcription, replication or reverse transcription
 CC with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as
 CC replication and transcription for their preparation. This polynucleotide
 CC sequence is a synthetic RNA 9A aptamer (100-mer) given in an

CC exemplification of the invention.
 XX Sequence 100 BP; 25 A; 22 C; 24 G; 0 T; 29 U; 0 Other;
 SQ Query Match 100.0%; Score 100; DB 12; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.4e-25;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGAGUGAGGAAUUCAGCAGGCAUAGUGCGACUCCGUCUCCUCAAACCGAUUAA 60
 DB 1 GGGAGUGAGGAAUUCAGCAGGCAUAGUGCGACUCCGUCUCCUCAAACCGAUUAA 60

QY 61 AUUGUUUAGCAUAGCCUAGCGACAGCAAGCUCUCCG 100
 DB 61 AUUGUUUAGCAUAGCCUAGCGACAGCAAGCUCUCCG 100

RESULT 3
 ADJ10064
 ID ADJ10064 standard; DNA; 100 BP.
 XX AC ADJ10064;
 XX 17-JUN-2004 (first entry)
 XX Synthetic RNA 9A (51y87) aptamer (100-mer).
 XX ss; unnatural base; 5'-substituted-2-oxo(1H) - pyridin-3-yl;
 XX photoreactive group; biotin; fluorescent molecule;
 XX 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.
 XX Synthetic.
 XX OS
 XX Key Location/Qualifiers
 XX modified_base 87
 XX /*tag= a
 XX /mod_base= OTHER
 XX /note= "OTHER= 51y"
 XX WO2004007713-A1.
 XX 22-JAN-2004.
 XX 28-FEB-2003; 2003WO-JP002342.
 XX 17-JUL-2002; 2002JP-00208568.
 XX (RIKE) RIKEN KK.
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX Hirao I, Yokoyama S, Hirao M, Mitsui T;
 XX WPI; 2004-122944/12.
 XX Nucleosides or nucleotides comprising novel base particularly 5-
 PT substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing
 PT functional nucleic acids as reagents or therapeutics in treating
 PT diseases.
 XX Disclosure; Fig 10; 78pp; Japanese.

XX This invention relates to nucleosides or nucleotides that a novel
 CC unnatural base, namely the 5'-substituted-2-oxo(1H) - pyridin-3-yl group
 CC as a base. Specifically, it refers to nucleic acid molecules integrated
 CC with a base substituted at the 5-position with either a photoreactive
 CC group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group,
 CC biotin or a derivative thereof or a fluorescent molecule selected from
 CC fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrhodamine or
 CC derivatives thereof. The present invention describes preparing nucleic
 CC acids by carrying out transcription, replication or reverse transcription
 CC with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as
 CC base which is applied as template for the integration of any of the
 CC nucleotides into its complementary position. Accordingly, the nucleosides
 CC and nucleotides are useful in developing functional nucleic acids
 CC including antisense DNAs and RNAs, ribozymes and aptamers as reagents or
 CC therapeutics in treating diseases by gene therapy. The produced nucleic
 CC acids are chemically stable with improved ease of amplification.
 CC replication and transcription for their preparation. This polynucleotide
 CC sequence is a synthetic RNA 9A aptamer (100-mer) given in an

[illegible]

RESULT 8
ADJ10066 standard; DNA; 147 BP.

ID ADJ10066 standard; DNA; 147 BP.
XX AC ADJ10066;
XX AC ADJ10066;
XX 17-JUN-2004 (first entry)
XX Synthetic RNA OC aptamer (100-mer).
DE ss; unnatural base; 5'-substituted-2-oxo(1H)-pyridin-3-yl;
XX photoactive group; biotin; fluorescent molecule;
KW 6-substituted-2-aminopurin-9-yl; aptamer; gene therapy.
KW
XX Synthetic.
OS
XX WO2004007713-A1.
XX PD 22-JAN-2004.
XX 28-FEB-2003; 2003WO-JP002342.
XX PF 17-JUL-2002; 2002JP-00208568.
XX PR (RIKEN) RIKEN KK.
PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX Hirao I, Yokoyama S, Hirao M, Mitsui T;
PI WPI; 2004-122944/12.
XX DR Nucleosides or nucleotides comprising novel base particularly 5-
XX substituted-2-oxo(1H)-pyridin-3-yl group as base, useful in developing
PT functional nucleic acids as reagents or therapeutics in treating
PT diseases.
XX PS Disclosure; Fig 10; 78pp; Japanese.
XX CC This invention relates to nucleosides or nucleotides that a novel
CC unnatural base, namely the 5'-substituted-2-oxo(1H)-pyridin-3-yl group
CC as a base. Specifically, it refers to nucleic acid molecules integrated
CC with a base substituted at the 5-position with either a photoreactive
CC group (e.g. iodine or bromine), an alkenyl, alkynyl or amino group,
CC biotin or a derivative thereof or a fluorescent molecule selected from
CC fluorescein, 6-carboxyfluorescein, tetramethyl-7-carboxyrhodamine or
CC derivatives thereof. The present invention describes preparing nucleic
CC acids by carrying out transcription, replication or reverse transcription
CC with a nucleic acid containing a 6-substituted-2-aminopurin-9-yl group as
CC base which is applied as template for the integration of any of the
CC nucleotides into its complementary position. Accordingly, the nucleosides
CC and nucleotides are useful in developing functional nucleic acids
CC including antisense DNAs and RNAs, ribozymes and aptamers as reagents or
CC therapeutics in treating diseases by gene therapy. The produced nucleic
CC acids are chemically stable with improved ease of amplification,
CC replication and transcription for their preparation. This polynucleotide
CC sequence is a synthetic RNA OC aptamer (100-mer) given in an
CC exemplification of the invention.
XX SQ Sequence 147 BP; 32 A; 34 C; 37 G; 0 T; 44 U; 0 Other;

Query Match 50.6%; Score 50.6; DB 12; Length 147;
Best Local Similarity 75.2%; Pred. No. 8e-08;
Matches 76; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

QY 1 GGAGUGGAGGAUUAUCAUGGAGCAUAUGGACUCCGUCUCCAAACCGAU-UAAUA 59
DB 1 GGAGUGGAGGAUUAUCAUGGAGCAUAUGGACUCCGUCUCCAAACCGAUUUUGAGCUGAUU 60

QY 60 AAUUGGUUUUAUGCAUUGCCUUAAGGACGACGAGCUUUGCC 100
DB 61 CACCUGUAAACGCAUUGCCUUAAGGACGACGAGCUUUGCC 101

RESULT 9
AAZ99076
ID AAZ99076 standard; RNA; 45 BP.

XX AC AAZ99076;
XX AC AAZ99076;
XX 21-JUN-2000 (first entry)
XX Ras target protein-RNA binding sequence #1 for generating primers.
DE Ras target protein; malignant tumour; signal transmission regulation;
KW cell proliferation; cell differentiation; aptamer; inflammation; ss.
KW
XX Homo sapiens.
OS
XX WO200009684-A1.
XX PD 24-FEB-2000.
XX 13-AUG-1999; 99WO-JP004399.
XX PR 14-AUG-1998; 98JP-00242596.
XX PR 24-NOV-1998; 98JP-00333284.
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX PI Yokoyama S, Hirao I, Sakamoto K;
XX WPI; 2000-224330/19.
XX Nucleic acid e.g. RNA aptamer capable of binding specifically to Ras
PT target protein like Raf-1, useful in drug compositions to treat and
PT diagnose malignant tumors and inflammation.
XX PS Disclosure; Page 52; 59pp; Japanese.
XX CC The invention relates to novel nucleic acids which bind specifically to
CC the target protein of Ras, e.g. Raf-1. RNA sequences (AAZ99076-299077)
CC represent the sequence of aptamer inserts that bind the Ras target
CC protein which are used to generate PCR primers. Aptamers based on these
CC sequences are useful in the treatment and diagnosis of malignant tumours
CC and inflammation. The nucleic acids can be used to formulate medicinal
CC compositions that are useful in the treatment of malignant tumours and
CC inflammation and for disease diagnosis by binding specifically to Ras
CC target protein and regulating transmission of signal causing
CC proliferation or differentiation of cells
XX SQ Sequence 45 BP; 11 A; 10 C; 7 G; 0 T; 17 U; 0 Other;

Query Match 45.0%; Score 45; DB 3; Length 45;
Best Local Similarity 100.0%; Pred. No. 5.3e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 AUGUCACUCCGUCUCCUUAACACGUAUAUAUUGGUUUUAG 71
DB 1 AUGUCACUCCGUCUCCUUAACACGUAUAUAUUGGUUUUAG 45

RESULT 10
AAC75113/C
ID AAC75113 standard; cDNA; 748 BP.

XX AC AAC75113;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF668 polynucleotide sequence SEQ ID NO:1335.
DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiproliferative; antiparkinsonian; neuroprotective;
KW anticovulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant vasotropic; antidiabetic;
KW

[illegible]

QY 18 UCGAGGCAUUGUGGACUCCGCUUCCUCAAACCGAUUAAUUGUUUAGCAUUG 77
Db 317 TTGGGTGCGACCTCGAGCCGCGGTCTTTCAGACCGTGTGATGAACGGTGTAGGTCATG 258
QY 78 CCUUGAGCGACGACGACG 94
Db 257 CCTGGGCACGGGAAGC 241

RESULT 12
ID AAS59583 standard; DNA; 15776 BP.
AC AAS59583;
XX
XX 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein encoding DNA #78.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant; ds.

XX Propionibacterium acnes.
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US012865.
XX
XX 21-APR-2000; 2000US-0199047P.
XX 02-JUN-2000; 2000US-0208841P.
XX 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Claim 1; SEQ ID NO 78; 1069pp; English.

XX Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
CC infections of bone, joints and the central nervous system, however it is
CC particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptides shown in AAU57347-AAU57508 and AAU67590-AAU67591. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published_pct_sequences

XX Sequence 15776 BP; 3393 A; 4931 C; 4483 G; 2969 T; 0 U; 0 Other;

Query Match 30.6%; Score 30.6; DB 4; Length 15776;
Best Local Similarity 46.8%; Pred. No. 4;
Matches 36; Conservative 12; Mismatches 29; Indels 0; Gaps 0;
QY 18 UCGAGGCAUUGUGGACUCCGCUUCCUCAAACCGAUUAAUUGUUUAGCAUUG 77
Db 8427 TTGGGTGCGACCTCGAGCCGCGGTCTTTCAGACCGTGTGATGAACGGTGTAGGTCATG 8486
QY 78 CCUUGAGCGACGACGACG 94
Db 8487 CCTGGGCACGGGAAGC 8503

RESULT 13
ID ACF64512 standard; DNA; 15776 BP.
XX
XX ACF64512;
XX
XX 17-OCT-2003 (first entry)
XX
DE Propionibacterium acnes DNA contig sequence #78.
XX
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine; ds.
XX
XX Propionibacterium acnes.
XX WO2003033515-A1.
XX 24-APR-2003.
XX
XX 11-OCT-2002; 2002WO-US032727.
XX
XX 15-OCT-2001; 2001US-00978825.
XX
XX (CORI-) CORIXA CORP.
XX
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Douglas J;
XX
XX WPI; 2003-381789/36.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.

XX Claim 1; SEQ ID NO 78; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
XX and the kit is useful for performing a diagnostic assay. The present

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2004, 11:12:16 ; Search time 51.1111 Seconds
(without alignments)
1390.674 Million cell updates/sec

Title: US-09-529-397C-25
Perfect score: 100
Sequence: 1 gggagggaggaaucaucg.....uagcgacagcaagcuucgc 100

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/prodata/1/ina/5A COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	30.4	30.4	366	4	US-09-328-352-608
2	29.8	29.8	4285	3	US-09-410-464-1
3	27.4	27.4	405	4	US-09-270-767-6441
4	27.4	27.4	405	4	US-09-270-767-21723
C 5	27.4	27.4	3791	3	US-08-675-773B-3
C 6	27.0	27.0	208	4	US-09-513-999C-21500
C 7	26.8	26.8	415	4	US-09-513-999C-26061
8	25.8	25.8	798	4	US-08-956-171E-162
9	25.6	25.6	372	4	US-08-956-171E-798
10	25.6	25.6	372	4	US-08-956-171E-798
11	25.6	25.6	372	4	US-08-781-986A-738
12	25.2	25.2	355	4	US-09-513-999C-8830
C 13	25.2	25.2	786431	4	US-09-751-389-3
14	25	25.0	246	4	US-09-513-999C-35958
C 15	25	25.0	960	4	US-09-328-352-1916
16	25	25.0	1830121	4	US-09-557-884-1
17	25	25.0	1830121	4	US-09-643-990A-1
18	25	25.0	1830121	4	US-10-329-960-1
19	24.8	24.8	2775	4	US-09-248-796A-4695
20	24.8	24.8	3786	3	US-08-975-762-42
21	24.8	24.8	3786	3	US-09-295-028-42
22	24.8	24.8	3786	3	US-09-106-582-42
23	24.8	24.8	3786	4	US-09-159-469-42
24	24.8	24.8	3786	4	US-09-693-542-42
25	24.8	24.8	7399	2	US-08-418-848A-9
26	24.8	24.8	9709	2	US-08-188-583-5
27	24.8	24.8	9709	3	US-08-388-353-1

28	24.8	24.8	9709	3	US-08-488-551B-1
29	24.8	24.8	9709	4	US-09-309-572-15
30	24.8	24.8	9709	4	US-09-718-096-15
31	24.8	24.8	12479	4	US-09-318-138-13
32	24.8	24.8	12494	3	US-08-935-312-13
33	24.8	24.8	12494	3	US-08-848-760B-33
34	24.8	24.8	15581	3	US-08-646-538-35
35	24.8	24.8	15581	3	US-09-503-222-35
C 36	24.6	24.6	544	4	US-09-621-976-1765
37	24.6	24.6	1212	4	US-09-248-796A-136
38	24.6	24.6	1827	4	US-09-270-767-29735
39	24.6	24.6	2612	3	US-09-105-390-7
40	24.6	24.6	4285	4	US-09-270-767-13714
41	24.6	24.6	11802	4	US-08-956-171E-70
42	24.6	24.6	11802	4	US-08-781-986A-70
C 43	24.4	24.4	915	4	US-09-919-497-19
44	24.4	24.4	13542	4	US-08-956-171E-154
45	24.4	24.4	13542	4	US-08-781-986A-154

ALIGNMENTS

RESULT 1
US-09-328-352-608/c
; Sequence 608, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 608
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-608

Query Match 30.4%; Score 30.4; DB 4; Length 366;
Best Local Similarity 39.8%; Pred. No. 0.088;
Matches 35; Conservative 17; Mismatches 36; Indels 0; Gaps 0;

QY	1	GGGAGUGAGGAUUCGAGGAGCAUUGCGUCCUCCUUAACCAAGUUAUA 60
Db	101	GGGAGTTTATGATATAGCTTTCTATATGTCGACCAGTAATTGTGCAACAGTAATAG 42
QY	61	AUUGGUUUAUGCAUUGCCUUAAGGACA 88
Db	41	TCTGGTTATCGGTATATTCTAGATTCA 14

RESULT 2
US-09-410-464-1
; Sequence 1, Application US/09410464
; Patent No. 6395892
; GENERAL INFORMATION:
; APPLICANT: Strauss et al.
; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
; FILE REFERENCE: poplar and other plant species.
; CURRENT APPLICATION NUMBER: 53375
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 09/287,700
; EARLIER FILING DATE: 1999-04-06
; EARLIER APPLICATION NUMBER: 60/080,851
; EARLIER FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4285

```
; TYPE: DNA
; ORGANISM: Populus balsamifera subsp. trichocarpa
US-09-410-464-1

Query Match      29.8%; Score 29.8; DB 3; Length 4285;
Best Local Similarity 33.3%; Pred. No. 0.33;
Matches 27; Conservative 22; Mismatches 32; Indels 0; Gaps 0;

Qy 8 GAGGAUUCAGCAGGCAUAGUCGACUCCGUCUCCUACCAACCAAGUUAUAAUUGGU 67
Db 1193 GATTATCTCTGAACCTTTCTGTTTTTTATGCTTCAATCCATATTATTATGTTT 1252

Qy 68 UAGCAUAGCCUUAAGGACA 88
Db 1253 TTATGATTTTCTTAGATACA 1273

RESULT 3
US-09-270-767-6441
; Sequence 6441, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6441
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-6441

Query Match      27.4%; Score 27.4; DB 4; Length 405;
Best Local Similarity 35.5%; Pred. No. 1.2;
Matches 33; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

Qy 7 GGAGGAUUCAGCAGGCAUAGUCGACUCCGUCUCCUACCAACCAAGUUAUAAUUGGU 66
Db 227 GGTGTAATCCATTGGTTCAATTTGGTCTCTCACTCAACCTGTAATAATCCAT 286

Qy 67 UUAGCAUAGCCUUAAGGACAAGCUUCUG 99
Db 287 TTTTAAAGTCGCTGAACCAACCGAGATCTG 319

RESULT 4
US-09-270-767-21723
; Sequence 21723, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 21723
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-21723

Query Match      27.4%; Score 27.4; DB 4; Length 405;
Best Local Similarity 35.5%; Pred. No. 1.2;
Matches 33; Conservative 19; Mismatches 41; Indels 0; Gaps 0;

Qy 7 GGAGGAUUCAGCAGGCAUAGUCGACUCCGUCUCCUACCAACCAAGUUAUAAUUGGU 66
Db 227 GGTGTAATCCATTGGTTCAATTTGGTCTCTCACTCAACCTGTAATAATCCAT 286

Qy 67 UUAGCAUAGCCUUAAGGACAAGCUUCUG 99
Db 287 TTTTAAAGTCGCTGAACCAACCGAGATCTG 319
```

```
Qy 67 UUAGCAUAGCCUUAAGGACAAGCUUCUG 99
Db 287 TTTTAAAGTCGCTGAACCAACCGAGATCTG 319

RESULT 5
US-08-675-773B-3/c
; Sequence 3, Application US/08675773B
; Patent No. 6166288
; GENERAL INFORMATION:
; APPLICANT: DIAMOND, LISA E
; APPLICANT: LOGAN, JOHN S
; APPLICANT: BYRNE, GUERARD W
; APPLICANT: SHARMA, AJAY
; TITLE OF INVENTION: METHOD OF PRODUCING TRANSGENIC ANIMALS
; TITLE OF INVENTION: FOR XENOTRANSPLANTATION. . .
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 SEVENTH STREET, N.W., SUITE 300
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/675,773B
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28,005
; REFERENCE/DOCKET NUMBER: DIAMOND-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-675-773B-3

Query Match      27.4%; Score 27.4; DB 3; Length 3791;
Best Local Similarity 41.6%; Pred. No. 2.5;
Matches 32; Conservative 14; Mismatches 31; Indels 0; Gaps 0;

Qy 19 CGAGGCAUAGUCGACUCCGUCUCCUACCAACCAAGUUAUAAUUGGUUUAUAGCAUAGC 78
Db 624 CCAGGATTCAGACAGCCCTGTCTCCCTGAACCCCTCTATTAGATGTTGTGAGCATTTTC 565

Qy 79 CUUAGCGCAGCAGCAGCU 95
Db 564 CATGGGACTTGATGCT 548

RESULT 6
US-09-513-999C-21500/c
; Sequence 21500, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
```



```
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 21500
LENGTH: 208
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13
OTHER INFORMATION: v=a or c or g
US-09-513-999C-21500

Query Match 27.0%; Score 27; DB 4; Length 208;
Best Local Similarity 39.8%; Pred. No. 1.3;
Matches 33; Conservative 15; Mismatches 35; Indels 0; Gaps 0;

QY 13 AUUACUGAGGCAUUGUCGACUCCGUCUCCUCAAACCAAGUAAUUGGUUUUAGC 72
Db 203 ATTAACACACCAACTGTCAAGTCTGCTTTAAAAAACCAGCTTGACATTTGCTTCATC 144

QY 73 AUAGCCUUAAGGCACAGCAAGCU 95
Db 143 ATATTCAGACCCCGCAGCTAGCT 121

RESULT 7
US-09-513-999C-26061/c
Sequence 26061, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 26061
LENGTH: 415
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 340
OTHER INFORMATION: y=c or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 344
OTHER INFORMATION: k=g or t
US-09-513-999C-26061

Query Match 26.8%; Score 26.8; DB 4; Length 415;
Best Local Similarity 48.4%; Pred. No. 2;
Matches 30; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

QY 28 UGUCGACUCGUCUCCUUAACACCAAGUAAUUGGUUUUAGCAUAGCCUUAAGCGAC 87
Db 269 TGTAGACACCACTGGCATTACACCACTAGTGGTGGTCCATATCCCAAAGAGAC 210

QY 88 AG 89
Db 209 AG 208

FILE REFERENCE: 59.US2.REG
CURRENT APPLICATION NUMBER: US/08/956,171E
Sequence 162, Application US/08956171E
Patent No. 6593114
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
Gil H. Choi
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1324
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 162:
US-08-956-171E-162

Query Match 25.8%; Score 25.8; DB 4; Length 798;
Best Local Similarity 31.2%; Pred. No. 5.7;
Matches 24; Conservative 21; Mismatches 32; Indels 0; Gaps 0;

QY 22 GGCAUAGUCGACUCCGUCUCCUCAAACCAAGUAAUUGGUUUUAGCAUAGCCUU 81
Db 196 GGCATCTTTCACCTTCTCTTATTTTCGTCAGTATTAAAGTGGTTTAGATTCCTCAT 255

QY 82 AGCGACAGCAAGCUUCU 98
Db 256 CGCTTCTGCGATTCTTCT 272

RESULT 9
US-08-781-986A-162
Sequence 162, Application US/08781986A
Patent No. 6737248
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
```

```
;
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-781-986A-162
;
; Query Match 25.8%; Score 25.8; DB 4; Length 798;
; Best Local Similarity 31.2%; Pred. No. 5.7;
; Matches 24; Conservative 21; Mismatches 32; Indels 0; Gaps 0;
;
; QY 22 GCGCAUGCGAGCUCGUCUCCUACAAACAGUAAUAAUUGUUUAGCAUAGCCUU 81
; Db 196 GGCATCTTTCACCTTCTTCTTATTTTCGTCAGTATTAAGTGTTTAGGATTCCTCAT 255
;
; QY 82 AGCGACAGCAAGCUUCU 98
; Db 256 CGCTTCGCGATTTCCT 272
;
; RESULT 10
; US-08-956-171E-798
; Sequence 798, Application US/08956171E
; Patent No. 6593114
;
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
;
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 798:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-956-171E-798
;
; Query Match 25.6%; Score 25.6; DB 4; Length 372;
; Best Local Similarity 36.9%; Pred. No. 5.3;
; Matches 24; Conservative 16; Mismatches 25; Indels 0; Gaps 0;
;
; QY 27 AUGUCGACUCGUCUCCUACAAACAGUAAUAAUUGUUUAGCAUAGCCUAGCGA 86
; Db 223 ATTCAACACCTTTTCTTTCATACCTTTTAAACAGGTTGTGTGCAATTTTTCGA 282
;
; QY 87 CAGCA 91
; Db 283 GGCCA 287
;
; RESULT 11
; US-08-781-986A-798
; Sequence 798, Application US/08781986A
; Patent No. 6737248
;
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 798:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-956-171E-798
;
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 798:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 372 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-956-171E-798
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-798

Query Match      25.6%; Score 25.6; DB 4; Length 372;
Best Local Similarity 36.9%; Pred. No. 5.3; Mismatches 25; Indels 0; Gaps 0;
Matches 24; Conservative 16;

QY 27 AUGUGACUCCGUCUCCUUAACACAGUUAUUAUUGGUUAGCAUAGCCUUAAGCGA 86
Db 223 ATTCAACACCTTTTCTTTTCATACCTTTTAAACAGGTTGTGTCATTGNTTTTTCGA 282

QY 87 CAGCA 91
Db 283 GGCCA 287

RESULT 12
US-09-513-999C-8830
; Sequence 8830, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 8830
; LENGTH: 355
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 309
; OTHER INFORMATION: k=g or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 319
; OTHER INFORMATION: r=a or g
US-09-513-999C-8830

Query Match      25.2%; Score 25.2; DB 4; Length 355;
Best Local Similarity 37.1%; Pred. No. 7.3; Mismatches 23; Indels 0; Gaps 0;
Matches 23; Conservative 16;

QY 27 AUGUGACUCCGUCUCCUUAACACAGUUAUUAUUGGUUAGCAUAGCCUUAAGCGA 86
Db 166 ATTCAACCTCTGTTTCTTAAACCGTTTTTCTTTGTTGTTGTCATTAGCATGTGCGA 225

QY 87 CA 88
Db 226 CA 227

RESULT 13
US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
```

```
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751.389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match      25.2%; Score 25.2; DB 4; Length 786431;
Best Local Similarity 42.6%; Pred. No. 91; Mismatches 13; Mismatches 18; Indels 0; Gaps 0;
Matches 23; Conservative 13;

QY 44 CUUCAACCAAGUUAUUAUUGGUUUAAGCAUAGCCUUAAGCGACAGCAAGCUUC 97
Db 507097 CTCAGAGCTGTGATATATTTTAAAGCCATTGACTTAGTGACAAATTTGTTTC 507044

RESULT 14
US-09-513-999C-35958
; Sequence 35958, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 35958
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 137
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: 143
; OTHER INFORMATION: r=a or g
; NAME/KEY: misc_feature
; LOCATION: 153
; OTHER INFORMATION: w=a or t
; NAME/KEY: misc_feature
; LOCATION: 170
; OTHER INFORMATION: b=c or g or t
US-09-513-999C-35958

Query Match      25.0%; Score 25; DB 4; Length 246;
Best Local Similarity 43.9%; Pred. No. 7.7; Mismatches 13; Mismatches 10; Indels 0; Gaps 0;
Matches 18; Conservative 13;

QY 30 UCGACUCCGUCUCCUUAACCAAGCAUUAUUAUUGGUUUUA 70
Db 182 TAGCTGGGCAATCTTCATGCCAGTAGTAATTTGGTTTA 222
```

```

RESULT 15
US-09-328-352-1916/c
; Sequence 1916, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1916
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1916

```

Query Match	25.0%	Score 25;	DB 4;	Length 960;
Best Local Similarity	36.9%	Prd. No. 12;		
Matches 24;	Conservative 16;	Mismatches 25;	Indels 0;	Gaps 0;
QY	12	AAUUCACUGAGGCAGUAUGUCACUCGCGUCUCCUUCACAAACGAGUUAUAUAUUGGUUUUAG	71	
DB	534	AATTGCTTAGCAATACGGCGACTATCGTGTCTTCCACCATATGATAAATGATTATTCG	475	
QY	72	CAUACU	76	
DB	474	CAAAAT	470	

Search completed: November 30, 2004, 11:53:59
Job time : 58.1111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 30, 2004, 11:12:15 ; Search time 231.667 Seconds
(without alignments)
2339.337 Million cell updates/sec

Title: US-09-529-397C-25

Perfect score: 100

Sequence: 1 gggaguggaggaaucaucg.....uagcagcagcaagcuucg 100

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3640242 seqs, 2709731945 residues

Total number of hits satisfying chosen parameters: 7280484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	29.8	29.8	4285	14	US-10-104-580-1	
C 2	28.8	28.8	675	16	US-10-282-122A-40631	
C 3	28.6	28.6	2000	9	US-09-938-842A-2908	
C 4	28.6	28.6	2000	11	US-09-938-842A-2908	
C 5	28.2	28.2	586	16	US-10-425-114-30801	
C 6	28.2	28.2	1612	17	US-10-767-701-13576	
C 7	28.2	28.2	1851	18	US-10-425-115-121623	
C 8	28.0	1691139	14	US-10-067-514-1	Sequence 1, Appl1	
C 9	28.0	1691139	16	US-10-419-723-1	Sequence 1, Appl1	
C 10	27.8	27.8	1093	15	US-10-369-493-25760	Sequence 25760, A
C 11	27.8	27.8	99291	17	US-10-322-281-744	Sequence 744, App
C 12	27.0	27.0	522	15	US-10-097-111-47	Sequence 47, Appl

13	27	56506	15	US-10-097-111-10	Sequence 10, Appl
14	26.8	1115	15	US-10-271-181B-6	Sequence 6, Appli
15	26.6	1884	15	US-10-369-493-35948	Sequence 35948, A
C 16	26.6	48995	16	US-10-052-482-85	Sequence 85, Appl
C 17	26.4	621	13	US-10-027-632-199104	Sequence 199104,
C 18	26.4	621	13	US-10-027-632-199104	Sequence 199104,
C 19	26.4	621	13	US-09-814-353-15351	Sequence 15351, A
C 20	26.2	502	13	US-10-027-632-324279	Sequence 324279,
C 21	26.2	502	13	US-10-027-632-324280	Sequence 324280,
C 22	26.2	502	15	US-10-027-632-324279	Sequence 324279,
C 23	26.2	502	15	US-10-027-632-324280	Sequence 324280,
C 24	26.2	181684	13	US-10-087-192-790	Sequence 790, App
C 25	26	368	16	US-10-424-599-130987	Sequence 130987,
26	26	725	13	US-10-027-632-150280	Sequence 150280,
27	26	725	13	US-10-027-632-150280	Sequence 150280,
28	26	740	13	US-10-027-632-14431	Sequence 14431, A
29	26	740	13	US-10-027-632-14432	Sequence 14432, A
30	26	740	15	US-10-027-632-14431	Sequence 14431, A
31	26	2515	13	US-10-027-632-14432	Sequence 14432, A
32	26	2515	13	US-10-027-632-101724	Sequence 101724,
33	26	2515	15	US-10-027-632-101724	Sequence 101724,
34	26	7285	14	US-10-198-846-13160	Sequence 13160, A
35	26	8998	15	US-10-311-455-1680	Sequence 1680, Ap
36	25.8	798	8	US-08-781-986A-162	Sequence 162, App
37	25.8	798	16	US-10-329-624-162	Sequence 162, App
C 38	25.8	1165	18	US-10-425-115-310	Sequence 310, App
C 39	25.8	1550	17	US-10-437-963-73812	Sequence 73812, A
C 40	25.8	2577	17	US-10-437-963-73810	Sequence 73810, A
C 41	25.8	2688	17	US-10-437-963-73813	Sequence 73813, A
42	25.8	25320	13	US-10-087-192-253	Sequence 253, App
43	25.8	43980	16	US-10-398-221-5	Sequence 5, Appli
44	25.8	94330	13	US-10-087-192-1222	Sequence 1222, Ap
C 45	25.8	226215	13	US-10-087-192-1948	Sequence 1948, Ap

ALIGNMENTS

RESULT 1

US-10-104-580-1
; Sequence 1, Application US/10104580

; Publication No. US20030033628A1

; GENERAL INFORMATION:

; APPLICANT: Strauss et al.

; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in poplar and other plant species.

; FILE REFERENCE: 62486

; CURRENT APPLICATION NUMBER: US/10/104,580

; CURRENT FILING DATE: 2002-03-21

; PRIOR FILING DATE: 1999-10-01, 464

; PRIOR APPLICATION NUMBER: 09/287,700

; PRIOR FILING DATE: 1999-04-06

; PRIOR APPLICATION NUMBER: 60/080,851

; PRIOR FILING DATE: 1998-04-06

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 4285

; TYPE: DNA

; ORGANISM: Populus balsamifera subsp. trichocarpa

US-10-104-580-1

Query Match 29.8%; Score 29.8; DB 14; Length 4285;
Best Local Similarity 33.3%; Pred. No. 9.7; Indels 0; Gaps 0;
Matches 27; Conservative 22; Mismatches 32

QY 8 GAGGAUUCUUGCAGGCAUUGUCACUCCGCUUCCAAACCAAGUUAUAAUUGUU 67
DB 1193 GATTTAATCCGTAACTTCTTCTTTTANGCCTTCAATCCATCTATTATTTT 1252

QY 68 UUAGCAUUGCCUAGCGACA 88


```
RESULT 5
US-10-425-114-30801/c
; Sequence 30801, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30801
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73080F09_FLI
US-10-425-114-30801

Query Match      28.2%; Score 28.2; DB 16; Length 586;
Best Local Similarity 37.1%; Pred. No. 17;
Matches 33; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY      8 GAGGAUUAUCCGAGCAUUGUGACUCCGUCUCCUACCAACAGUUAUAAUUGGU 67
Db      92 GGGGCATTTTATGTTGTTATTAGGATTTCTCTGCTCCACCGCATCACTCGGTT 33

QY      68 UUAGCAUUAUCCUUAAGCGACAGCAAGCU 96
Db      32 CCAGCAAGCGCATTAGGGAAGGAAGATT 4

RESULT 6
US-10-767-701-13576/c
; Sequence 13576, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13576
; LENGTH: 1612
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS1303_1
US-10-767-701-13576

Query Match      28.2%; Score 28.2; DB 17; Length 1612;
Best Local Similarity 38.2%; Pred. No. 24;
Matches 34; Conservative 17; Mismatches 38; Indels 0; Gaps 0;

QY      8 GAGGAUUAUCCGAGCAUUGUGACUCCGUCUCCUACCAACAGUUAUAAUUGGU 67
Db      1152 GGGGCATTTTATGTTGTTATTAGGATTTCTCTGCTCCACCGCATCACTCGGTT 1093

QY      68 UUAGCAUUAUCCUUAAGCGACAGCAAGCU 96
Db      1092 CCAGCAAGCGCATTAGGGAAGGAAGATT 1064
```

```
RESULT 7
US-10-425-115-121623/c
; Sequence 121623, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 121623
; LENGTH: 1851
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_42398C.1
US-10-425-115-121623

Query Match      28.2%; Score 28.2; DB 18; Length 1851;
Best Local Similarity 37.1%; Pred. No. 26;
Matches 33; Conservative 18; Mismatches 38; Indels 0; Gaps 0;

QY      8 GAGGAUUAUCCGAGCAUUGUGACUCCGUCUCCUACCAACAGUUAUAAUUGGU 67
Db      1333 GGGGCATTTTATGTTGTTATTAGGATTTCTCTGCTCCACCGCATCACTCGGTT 1274

QY      68 UUAGCAUUAUCCUUAAGCGACAGCAAGCU 96
Db      1273 CCAGCAAGCGCATTAGGGAAGGAAGATT 1245

RESULT 8
US-10-067-514-1/c
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match      28.0%; Score 28; DB 14; Length 1691139;
Best Local Similarity 41.7%; Pred. No. 3.6e+02;
Matches 25; Conservative 15; Mismatches 20; Indels 0; Gaps 0;

QY      12 AAUUAUCCGAGCAUUGUGACUCCGUCUCCUACCAACAGUUAUAAUUGGUUAG 71
Db      127542 AGTTGATCCATGCTTTTGAGGGAGCTTATCCCTTCAAAGCAGTAATAAAGTCTTTAG 127483

RESULT 9
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
```


[illegible]

```

RESULT 14
US-10-271-181B-6
; Sequence 6, Application US/10271181B
; Publication No. US20030158672A1
; GENERAL INFORMATION:
; APPLICANT: Kalyanaram Ramnarayan
; APPLICANT: Edward T. Maggio
; APPLICANT: P. Patrick Hess
; TITLE OF INVENTION: Use of Computationally Derived Protein
; TITLE OF INVENTION: Structures of Genetic Polymorphisms in Pharmacogenomics for
; TITLE OF INVENTION: Drug Design and Clinical Applications
; FILE REFERENCE: 24737-1906D
; CURRENT APPLICATION NUMBER: US/10/271,181B
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: 09/438,566
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: 09/704,362
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/709,905
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Human Immunodeficiency Virus (HIV)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(297)
; OTHER INFORMATION: HIV Protease
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (298)...(1116)
; OTHER INFORMATION: Portion of HIV Reverse Transcriptase
US-10-271-181B-6

```

```
Query Match      26.8%; Score 26.9; DB 15; Length 1116;
Best Local Similarity 38.4%; Pred. No. 65;
Matches 33; Conservative 16; Mismatches 37; Indels 0; Gaps 0;
```

```

Db      800  TAGAGCCTTCAGAAACAAATCCAGACATGCTCATCTATCAATACATGATGATTGT 859
Qy      66  UUUUAGCAUAUGCCUUAUGCACAGCA 91
      : : : : : : : : : : : : : :
Db      860  ATCTAGGATCTGACTTAGAAATAGGA 885

RESULT 15
US-10-369-493-35948
; Sequence 35948, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPANSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10 (52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35948
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
US-10-369-493-35948

Query Match      26.6%; Score 26.6; DB 15; Length 1884;
Best Local Similarity 49.0%; Pred. No. 94;
Matches 24; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

Qy      26  UAUGUGCAGCUCGUCUUCUCAAACAGUAUAAUUGUUUUUAGCAU 74
      : : : : : : : : : : : : : : : : : :
Db      364  TATGGCAGCGCGTCGGCCCATCAACCGGCTATGATTGGCTCGCCT 412

Search completed: November 30, 2004, 12:01:02
Job time : 237.667 secs

```

